Data Science for Business Analytics and Business Intelligence

3

Matthias Lederer
International School of Management (ISM), Munich, Germany

Patrick Schmid
International School of Management (ISM), Munich, Germany

INTRODUCTION

In today's globalized world, companies are no longer competing solely on the basis of the classic production factors such as capital, land and labor. Competitive advantages of an organization are increasingly relying on data or further processed to knowledge (Woodside and Quaddus, 2015). Prominent examples of this shift are the digital players:

- Facebook uses the data to promote advertising campaigns more targeted than any other media company before.
- Amazon collects data from many areas of everyday life such as purchases and products (Amazon Shop), entertainment (Amazon Echo), external services (Amazon WebServices, AWS) and more recently health information (Amazon wants to venture into health services (Isidore, 2018)).

The combination of data results in a variety of application scenarios and new business models (McCallum and Gleason, 2013). They often have the potential to disrupt entire markets (Aluya, 2014). At the same time, many small and specific use cases exist in which companies use innovative methods of data collection and analysis:

- Millions of DNA samples are compared to detect the risk of cancer.
- Virtual agents (e.g., chatbots) can handle the semantics and syntax of input information and develop answers based on documents.

In summary, all of these examples use many digital data sources, process them and open up new information or even knowledge that is digitally available (Williams, 2017). The self-optimizing methods, tools and underlying techniques are typically subsumed under the term "Data Science" (McCallum and Gleason, 2013).

Applications of machine learning with corresponding predictions permeate nearly all areas (examples given with respective companies or institutions) (Siegel, 2013):

- Social (family and personal life): Prediction of future locations from position tracking (Nokia) and prediction of love from online messages (OkCupid).
- Marketing and advertising: Prediction of product choices from past purchases (Tesco) and prediction of cancellations (FedEx).

DOI: 10.4018/978-1-7998-3473-1.ch037

- Financial and insurance: Algorithmic trading (London Stock Exchange) and prediction of loan defaults from past data (Citigroup).
- Healthcare: Prediction of clinical trial recruitment (GSK) and prediction of billing errors (Washington state hospitals).
- Fraud detection: Prediction of tax returns (IRS) and prediction of worker's compensation claims (U.S. Postal Service).
- Fault detection for efficiency and safety: Prediction of oil flow rate via neural networks (National Iranian South Oil Company) and prediction of travel time via traffic analysis (New South Wales, Australia).
- Human resources: Prediction of quitting (Hewlett-Packard) and prediction of job performance from fitness data (U.S. Special Forces).
- Public administration and education: Prediction of student performance (IBM) and prediction of student grading to enable automated grading (Hewlett-Packard).
- Psychology: Prediction of dissatisfaction (Citibank) from the use of certain words.

BACKGROUND

As it might already have become apparent from the introduction, a lot of benefits can be created from the application of data science. But this application requires a somewhat new approach which is connected very much to the idea of agile project management (of data science projects). All the examples given so far are assessed in greater depth by Cao via naming relevant success factors, in particular:

- appropriate thinking approach
- adequate assessment of the complexity of the problem
- adequate modeling
- evaluation of model outcomes
- business engagement.

Furthermore, a distinction between general vs. domain-specific algorithms as well as a distinction between vendor-dependent vs. independent solutions is helpful (chapter 9 of Cao, 2018).

FOCUS OF THE ARTICLE

After specific examples and their background have now been mentioned, this chapter explains stepwise and in much detail how nowadays the challenge of harnessing the abundant amount of existing data is addressed effectively. Thereby it adheres to the usual differentiation between business intelligence and business analytics. For business intelligence it describes in particular the generic process of modern data organization that most data science applications follow, whereas for data analytics it places emphasis on showing that there exist a lot of different approaches which can be used.